DOCUMENT RESUME

ED 036 398 24 RE 002 353

AUTHOR MOSBERG, LUDWIG; SHIMA, FRED
TITLE COMPREHENSION OF CONNECTED DISCOURSE.
INSTITUTION SOUTHWEST REGIONAL EDUCATIONAL LAB., INGLEWOOD, CALIF.
SPONS AGENCY OFFICE OF EDUCATION (DHEW), WASHINGTON, D.C. BUREAU OF RESEARCH.
REPORT NO TR-12
BUREAU NO BR-6-2865
PUB DATE 1 APR 69
CONTRACT OEC-4-7-062865-3073
NOTE 41P.

EDRS PRICE EDRS PRICE MF-$0.25 HC-$2.15
DESCRIPTORS ASSOCIATIVE LEARNING, *INFORMATION PROCESSING, *INFORMATION THEORY, MEMORY, PAIRED ASSOCIATE LEARNING, *READING COMPREHENSION, *READING DEVELOPMENT, TIME FACTORS (LEARNING), *VERBAL LEARNING

ABSTRACT

A RATIONALE WAS DEVELOPED FOR RESEARCHING READING COMPREHENSION BASED ON INFORMATION GAIN. PREVIOUS DEFINITIONS OF COMPREHENSION WHICH WERE REVIEWED INCLUDED OPERATIONAL VS. NONOPERATIONAL AND SKILLS VS. PROCESSES. COMPREHENSION WAS VIEWED AS AN INFORMATIONAL PROCESSING EVENT WHICH INCLUDES A CONSTELLATION OF COGNITIVE AND LEARNING PROCESSES. TWO ARGUMENTS FOR THIS VIEW WERE MADE: (1) COMPREHENSION INSTRUCTION CAN BEST BE DEVISED THROUGH AN ANALYSIS OF THE COMPONENT PROCESSES INVOLVED IN THE EXTRACTION AND RECALL OF FACTUAL AND RELATIONAL INFORMATION IN READING MATERIAL AND (2) INSTRUCTION TO INCREASE COMPREHENSION ABILITY CAN BE PREDICATED UPON SUCH AN ANALYSIS. INFORMATION GAIN WAS DEFINED AS THE EXTRACTION AND RECALL OF NEW INFORMATION FROM A LANGUAGE STIMULUS AND WAS PURPOSED AS A MEASURE OF COMPREHENSION. DISCUSSION FOCUSED ON TYPES OF INFORMATION, INFORMATION GAIN, STIMULUS AND RESPONSE VARIABLES WHICH MAY AFFECT INFORMATION PROCESSING, AND STUDENT VARIABLES. DISCUSSION OF THEORETICAL ISSUES UNDERLYING INFORMATION PROCESSES INCLUDED LONG-TERM AND SHORT-TERM MEMORY, PROACTIVE AND RETROACTIVE INHIBITION, AND UNITS OF INFORMATION PROCESSING. AN EXTENSIVE BIBLIOGRAPHY IS INCLUDED. (WE)

RE

BR 6-2865 TR 12
PA-24
OE/BR

SOUTHWEST REGIONAL LABORATORY FOR EDUCATIONAL RESEARCH & DEVELOPMENT

ED 036398

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED
BY Southwest Regional Laboratory
TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE U.S. OFFICE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

RE 002 353

Comprehension of Connected Discourse

TR 12 1 APRIL 1969

© 1968 SOUTHWEST REGIONAL LABORATORY for Educational Research and Development, 11300 La Cienega Blvd., Inglewood, Calif., 90304
Published by Southwest Regional Laboratory for Educational Research and Development, a public agency supported as a regional educational laboratory by funds from the United States Office of Education, Department of Health, Education, and Welfare. The opinions expressed in this publication do not necessarily reflect the position or policy of the Office of Education, and no official endorsement by the Office of Education should be inferred.

ED036398

COMPREHENSION OF CONNECTED DISCOURSE

Ludwig Mosberg and Fred Shima[1]

I. INTRODUCTION

The reading process can be conceptualized as consisting of two components, identification and comprehension. The identification process is the recognition of graphemic symbols and their relation to the spoken language. Comprehension consists of the extracting, recalling and evaluating of information or meaning from the language stimulus (Carroll, 1964; Davis, 1956, 1966; Nordberg, 1956; Walcutt, 1967; Wiener & Cromer, 1967). Although there are still wide gaps in our understanding of identification skills and optimal methods of teaching them, the nature of the identification process is well defined. On the other hand, the subject-matter of comprehension remains vaguely defined. The present paper will examine and explicate the processes involved in comprehension.

---

[1] Grateful acknowledgement is made to Morton P. Friedman and Joseph F. Follettie for their constructive comments and editing of this paper.

Reading instruction typically proceeds by teaching identification skills using a phonics method, a whole-word method, or a combination of the two. Instruction in comprehension typically follows the acquisition of identification skills. This instruction is typically more implicit than explict. Usually the instructor probes the students for statements of main idea, evaluation statements of what is read, and other aspects often included under the rubric of "critical reading skills" (Niles, 1963; Simmons, 1965). Implicit is the notion that practice in answering different types of questions will somehow improve the student's ability to comprehend reading material. The rationale for the various components of instruction in comprehension is never explicitly stated, nor are the procedures for instruction or the criteria for measuring improvement.

The approach to comprehension taken in this paper is a rather radical departure from the above. Comprehension here is viewed as an information processing event which includes a constellation of cognitive and learning processes which interact in specified ways. Two arguments will be made: (1) that comprehension instruction can best be devised through a careful and systematic analysis of the component processes involved in the extraction and recall of factual and relational information in reading material, and (2) that instruction to increase comprehension ability can be predicated upon such analysis.

First, definitions of comprehension will be reviewed and a definition to guide the present analysis will be specified.

Definitions of Comprehension

One may distinguish two types of definitions of comprehension: (a) operational vs. non-operational and (b) skills vs. processes.

Operational definitions in terms of skills. This type of definition is best exemplified in the work of Davis (1944) and Holmes (1960, 1962, 1965). Both investigators use the technique of factor analysis to define the construct. They begin with a relatively large number of tests which purport to measure various aspects of comprehension, and by factor analysis attempt to determine the relationship between tests and to extract as many factors as the intercorrelations permit. These factors are generally named on the basis of what the tests which load on the factor have in common. These factors then define the construct of comprehension. The problem with this type of definition is that it is completely dependent upon the specific tests included in the factor matrix and that the obtained correlations will depend heavily on the reliability and validity of the measures. Such a definition begs the question, for it assumes that the tests used measure some aspect of comprehension, which further assumes that, implicity, at least, the construct is already defined.

Non-operational definitions in terms of skills. A host of other investigators have attempted to define comprehension in terms of a set of skills and abilities. These investigators have simply listed the skills believed involved without giving the rationale or the specific means for measuring them. This is perhaps the most common type of definition found in the literature. A partial list of the skills and abilities subsumed under the heading of comprehension is given below. The expanse of skills included in the construct of comprehension has resulted in it becoming an all encompassing "waste-basket construct."

Identifying main idea (Staiger & Bliesmer, 1956)
Recognition of fact and opinion (Betts, 1956; Davis, 1956)
Evaluating relevance of statements (Betts, 1956; Gans, 1940)
Drawing conclusions (Betts, 1956; McKee, 1948, Simmons, 1965)
Use of inference (Betts, 1956; Davis, 1956; Bedell, 1934)
Reasoning by analogy (Betts, 1956)
Organizing ideas (Betts, 1956; Langsman, 1941)
Following structure of passage (Davis, 1956)
Identifying tone and mood of passage (Davis, 1956)
Induction (Jordan, 1967)
Deduction (Jordan, 1967)
Whole-part recognition (Jordan, 1967)
Categorization (Jordan, 1967)
Use of previous learning (Niles, 1963)
Find and understand thought relationships (Niles, 1963)
Set specific purpose of reading (Niles, 1963)
Reflection (Park, 1966)
Going from literal to implied meaning (Robinson, 1958; Simmons, 1965)
Maintainence of interest (Robinson, 1958)
Use of context (Robinson, 1966)
Set proper rate (Robinson, 1966)
Assimilation and accommodation (Stauffer, 1967)
Noting detail (Traxler, 1951)

This list is not meant to be exhaustive but rather to indicate the range of skills incorporated by numerous authors under the heading of comprehension. Traxler (1951) analyzed 28 reading comprehension tests and found 49 types of reading skills supposedly tested. It is clear that these skills are not all independent, many most likely reflecting merely semantic differences, or differences in emphasis.

Definitions in terms of process. Definitions of comprehension in terms of process fall into two categories: (a) definitions given in terms of cognitive processes, higher-order mental processes, and thinking processes, and (b) definitions given in terms of information processing and communications systems.

A. Definitions in terms of cognitive processes. Carroll (1964) defines comprehension as a linguistic process of comprehending morphemes and grammatical constructions in which the morphemes occur. The lexical

meanings of morphemes, he suggests, can be described in terms of the objective referents, attributes and relationships. Meanings of grammatical constructions can be described in terms of the structural relationship among person, things and events in spatial-temporal configuration. He further holds that problems in comprehension result when the text contains lexical, grammatical or ideational materials which are not in the reader's repertoire. Carroll further assumes that comprehension occurs in response to some kind of internal representation of a spoken message. This definition suggests (a) that the study of reading comprehension should proceed to studying comprehension of spoken messages and the processes by which a written message is reconstructed in terms of the spoken message, and (b) that the study of comprehension should concern itself with both the semantic and grammatical components of messages. This definition is consistent with linguistic research presently underway at Southwest Regional Laboratory.

Gray (1960) identifies four processes in reading: (a) word perception, (b) comprehension, (c) reaction to what is read (critical reading), and (d) assimilation of new ideas with previous knowledge. Of interest here is the fact that Gray, unlike other authors, separates comprehension from reaction and assimilation. He defines comprehension on three levels, (a) comprehension of the literal meaning, i.e., what the author said, (b) comprehension of the implied meaning, i.e., what the author meant by the sequence of words used, and (c) the significance of the message. He has further argued (1951), quoting Thorndike (1917), that comprehension is a higher-order thinking process which involves: (1) "weighing of each of many elements in a sentence," (2) "their organization in their proper relations one to another," and (3) "the selection of certain of their connotations and rejection of others." Robinson (1966) has expanded Gray's model, but the processes have not yet been delimited in a manner which would lend itself to careful empirical study.

Stauffer (1967) defines comprehension in terms of two cognitive processes, assimilation and accommodation. By assimilation is meant the taking in and incorporating of what is perceived in terms of what is known and understood at the time. Accommodation refers to reorganization of conceptual structures until they fit and account for the new circumstances. As with other definitions, this one fails to specify how these processes are learned or developed or how they operate, i.e., what psychological processes are involved or how these processes may be studied empirically. Finally, he gives no explicit definition of what is meant by a conceptual structure.

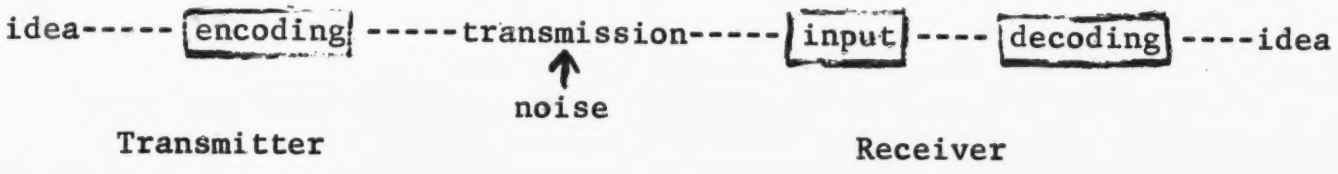
Johnson (1949) argues that reading is not a subject but rather a complex process which develops over the life-span of the individual. What this process consists of, however, is not dealt with except in terms of certain skills which the author claims are involved in, or are manifestations of, the process. In reading this definition, or any other so far considered, one asks himself questions: What makes this definition any more useful than any other? What does this definition add that others do not? The answers are typically negative.

Spache (1962), after criticizing Holmes' substrata theory of reading, attempts to define comprehension in terms of four processes: (a) cognition, (b) memory, (c) inductive reasoning, and (d) deductive reasoning. He argues that while skills and factors involved in comprehension can be identified, comprehension is a "gestalt" in which the sum is greater than its parts. While it is no doubt true that analysis of skills and abilities will not account for all the variance in the system, we take issue with the notion of invoking "gestalt" as an explanatory or helpful construct. By positing that the whole is greater than the parts one is merely admitting that he has neither identified the relevant factors or processes nor the proper relationships between those he has identified. The gestalt notion tends to remove the phenomenon from analytic investigation, which the present investigators are not prepared to do at this time. However, if by a "gestalt" Spache means that the comprehension process is a complex phenomenon with complex relationships existing between factors or processes, then, of course, we are in complete agreement. It is this latter notion of a "gestalt" which it is assumed Spache is advocating. The four processes which Spache considers to be involved in the comprehension of language need some elaboration. The processes he identifies are in terms suggested by Smith (1960) and Guilford (1960). Cognition refers to recognition of words by form, shape, structural parts, and context Memory refers to the recall of one of several associations to each word which is appropriate in the particular context. In comprehending a sentence, a chain of associations is elicited, and these associations in turn form higher-order associations on the basis of word groupings. These groupings of associations coalesce into the stated or implied meaning of a sentence. The meanings of successive sentences are inductively combined into the main idea. Furthermore, the sentence meanings may form the basis for various deductions. In a sense, Spache appears to be advocating a problem-solving approach to comprehension not unlike Gray's model.

B. Definitions in terms of information processing and communication systems. Kingston (1961, 1962) and Cleland (1965) define reading as a process of communications in which a message is transmitted, in a graphic mode, between individuals. For such communication to occur it is a necessary but not sufficient condition that the transmitter and receiver of the message agree on the meanings of the symbols employed. Kingston assumes that at least for the mature reader (and we find no reason to believe not for the "immature reader") each symbol elicits a set of responses, "depending upon his needs at the moment and the strength and number of his associations." Comprehension is the degree to which the reader's interpretation (elicited association in decoding) is congruent with that of the writer, or in absence of the writer, with some authority figure. As such, Kingston suggests, what is typically measured is conformity to the authority rather than comprehension. Kingston, then, sees the comprehension process as being a function of the congruence of associations to a given symbol between the transmitter (or authority) and the receiver, the familiarity of the reader

with the structural form of the message, and the comparability of the cognitive level of abstraction of the message with that of the reader. This model suggests that if a word association test were given to a group of Ss and if those Ss could be matched in terms of the similarity in protocols, then comprehension should be higher when a matched S writes a message than when a nonmatched S writes a message. We have not found any test of this hypothesis in the literature.

McCreary and Surkan (1965) attempt to expand on Kingston's model. The comprehension process is seen as a communication system channel which may be represented by the following schematic diagram based on the Shannon and Weaver (1949) model.



McCreary and Surkan point out that their model for reading is an analogy to hardware systems of communications and do not necessarily mean that the communication system in humans is adequately described in these terms. They do believe, however, that such an analogy provides a useful tool for the study of the processes involved in reading. The authors posit that the message in human communication is received "if some change is made on the mind of the reader or learner." The purpose of the message is to provide information to the receiver which helps him select some recognizable changes in the mind "from the spectrum of all those which could be chosen." The information extracted from a message depends on the number of possibilities that could be anticipated. This coincides with Shannon's (1948) suggestion that the amount of information in a message depends upon the number of alternative possibilities which the message eliminates. McCreary and Surkan attempt to describe the processes involved in the communication channel (the reader). During the processing of the message, i.e., prior to storage, five processes are identified: sampling, filtering, coding, decoding, and matching. The incoming message is sampled, irrelevant information and noise are filtered out, the message is coded in some language form, the information for retention is again filtered, the message is decoded or interpreted, and, finally, the message is matched with the receiver's prior knowledge and concepts. Storage of information is seen as occuring by a "chunking" process (Miller, 1956), but they make no assumptions about or distinctions between short-term and long-term memory.

The primary problem of conceptualizing the comprehension process in terms of a communication channel system is how "information" is to be defined. McCreary and Surkan define information in the tradition of information theory, i.e., the reduction of uncertainty. In order

to determine the uncertainty reduction of a message it would be necessary to know that probability of occurrence of each word in the passage as a function of the preceding words. This in itself would be a monumental task as Weaver and Weaver (1965) point out, but, in addition, this defintion of information is devoid of any notion of meaning. Given that the sequential probabilities of words were known it would then be possible to construct messages which would have a great amount of information in terms of entropy but which would be devoid of semantic meaning, and, in one sense at least, contain no "information" at all for the receiver of the message. Furthermore, as Semmelroth (1965) points out, since information varies with the number of alternatives, those alternatives must be assumed to be in the repertoire of the S. For alphabetic redundancy this is probably a safe assumption, but for lexical redundancy, particulary with children, this assumption is much less tenable. What is needed for the construction of a human information processing system is a definition of information which includes both the concept of information in terms of uncertainty in the statistical sense and of information in terms of the value of the semantic information.

It is painfully clear to the researcher that a wide variety of behaviors and processes are attributed to comprehension. Based on present knowledge it is not felt that it is advantageous, in terms of a systematic research approach, to define the construct of comprehension in an all-inclusive manner. Rather, attempts will be made to define various sub-components, to devise valid measures of each component process, to investigate the underlying processes involved, and to model these processes. In other words, comprehension will be viewed not as a unitary process, but rather as a system of processes involving linguistic, psychological and perceptual processes.

Initially, the defintion of comprehension will be restricted to the extraction and recall of new information from a language stimulus, which is regarded here as the primary function of reading (particulary within educational curriculum). At this time no attempt will be made to define "information" in any rigorous sense, or to consider the question of how semantic information in a language stimulus is to be quantified. This paper's main goal is the development of a comprehension curriculum. To this end, an attempt will be made to differentiate among the various kinds of information which are of research interest, and to investigate the processes involved in the comprehension of such information.

II. TYPES OF INFORMATION GAIN

"Information ga is used here as an index of comprehension. The reading of connected discourse may yield information of two general types: (a) word-for-word verbatim learning, and (b) substance learning. Included under verbatim learning are rote serial learning of part or

all of the passage, and learning of specifically stated facts in words taken directly from the text. Substance information is somewhat more difficult to define. Substance learning requires organization, interpretation and paraphrase of information. As such, comprehension may require the reader to deduce, induce and assimilate the information while processing the relationships between facts.

The primary interest here is in substance information gain, but verbatim recall will be discussed, for two reasons.

First, verbatim recall, particularly of stated facts, is often the transmitter's goal. Thus one might use a single passage to convey, as unrelated facts bearing on the topic "California," that its capitol is W, its population is X, its area is Y, and its principal industry is Z. Thus, California has attributes W,X,Y, and Z. Relations holding between attributes are not of interest. The goal of the message is simply to transmit a set of facts.

Second, verbatim recall permits objective measurement. Should verbatim and substance recall prove highly correlated, it might be possible to substitute verbatim measures for presently more subjectively defined measures of substance recall. A review of the experimental literature, however, does not support the view that verbatim recall measures can be used to predict substance recall.

It should be noted that literature on the verbatim-substance question is sparse and that most of it is pre-1940 work. Also, very few studies comparing verbatim and substance recall have used the same reading materials and/or the same sample of subjects. Many studies have obvious weaknesses. For example, Jones & English (1926) compared verbatim and substance recall using a different dependent variable for each: number of trials to a verbatim learning criterion vs. number of substance idea units correct. Trow (1928) examined both types of learning but tested only two subjects per condition. Several well-done studies, however, are available.

English, Welborn & Killian (1934) presented 1500-2500-word prose passages on psychology topics to college undergraduates. A true-false recognition test was given, with some items taken directly from the text and others being paraphrases or summaries of passage material. Scores on the verbatim items were higher than the substance items.

Cofer (1941) reported different results with shorter prose passages, 25-150 words, on Indian folk tales administered to college students. Verbatim information gain was measured by the number of words and sentences recalled from the text, and substance information gain was measured by the number of passage ideas recalled. More learning trials were required for verbatim than for substance learning. Thus verbatim learning was easier in the English, et al., study but more difficult in the Cofer study. This may be because a recognition test was used in the former study and a recall test in the latter.

More recently, Vernon (1951) investigated recall of discrete items and key general statements from 400-word passages on demographic topics administered to high school students. Specific recall was about 45% for each of the two test passages, but 59% of the substance information was recalled in one passage and only 30% in the other. Vernon concluded that verbatim learning does not necessarily ensure substance learning.

Yavuz (1963) reported paired-associates results relevant to the verbatim-substance learning distinction. Turkish-word stimuli and English-word responses were paired for training, and a retention test was given a week later. Although many of the correct responses were forgotten, most of the incorrect responses given by subjects had semantic ratings similar to those for the correct responses. In other words, the verbatim labels were missing but the substance content remained.

Sachs (1967) compared recognition memory of sentences which either were identical to those previously presented in passages or had been changed semantically or structurally. She found that as interpolated material between passage presentation and testing increased (from 0 to 80 to 160 syllables), recognition of semantic changes decreased only slightly while recognition of structural changes decreased greatly. These results indicate that substance gain, the meaning of sentences, is retained much better than verbatim gain, the original sequence of words.

The last two studies by Yavuz and Sachs suggest that psychological measures of meaning might be another way of investigating substance learning. If substance learning is "getting the meaning" of a message, and if meaning can be studied through word associations (Bousfield, 1961; Deese, 1966) and connotative ratings (Osgood, Suci & Tannenbaum, 1957; Osgood, 1966), then substance learning might be viewed as giving appropriate word associations and connotative ratings following presentation of a test passage. In word association tasks, the set of associations which particular subjects give in response to a stimulus word defines the meaning of the stimulus word. In tasks designed to evaluate connotative meaning, subjects rate a word on a battery of semantic differential scales; the profile of the scale ratings then defines the word. Semantic differential scales are 7-point scales anchored by bipolar adjectives, like Good-Bad, Fast-Slow, and Strong-Weak.

One problem is that tasks involving associative and connotative meaning typically deal with meanings of single words, whereas connected discourse covers meanings of word sequences. Responses to a word in context with other words differ from those given to words alone, e.g., LADY alone vs. BAD LADY.

Given that the student reads a test passage and then gives word associations or connotative ratings considered appropriate, does that signify that he acquired (comprehended) the meaning of the passage?

Essentially, yes. In an elemental sense, the substance underlying a highly ordered set of sentences has been transmitted. Associative and connotative indices do seem to get at the "bedrock" of meaning, but they offer only a diffuse set of impressions. What is needed is the demonstration of operations on which associations and connotations underlying comprehension are based. It is assumed that the passage material is remembered in some form and organized before the desired associations and connotations are drawn from it. Forms of retention and organization should be made explicit. Recall and recognition measures of substance learning reflecting forms of retention and organization should prove more relevant indices of substance information gain than discrete associations or ratings. This does not preclude using methods which assess meaning of single words when studying connected discourse. If, for example, embedding highly associated words in a passage leads to higher retention because the words act as mnemonic aids in storing and organizing the passage material, a test for single-word meaning may be appropriate.

Since a kind of essential retention of the original material is a part of substance learning, it is important to distinguish between verbatim and substance retention. Following is a sample test passage and sample test items to further clarify the difference between verbatim and substance learning.

> The Crusades were military expeditions undertaken by the Christian powers in the 11th, 12th and 13th centuries to recover the Holy Land from the Muslims. The First Crusade was in 1095, the Second in 1146, the Third in 1189, and the Fourth in 1200-1204. Although the First Crusade was successful, Saladin recaptured Jerusalem from the Christians in 1187 and maintained possession through the Third and Fourth Crusades.

In free recall, verbatim measures would be total number of words, content words (nouns, pronouns, verbs, adjectives, adverbs), and sentences identical to the original text. Substance indices would include synonyms of original text words and number of idea units. Passages would be scaled by judges beforehand into idea units which convey the sense of a phrase without requiring the original words. A possible breakdown into idea units might be: "The Crusades/ were military expeditions/ undertaken by the Christian powers/ and so on."

A modified cloze test procedure can also be used to compare the two types of learning. One exposure of the passage with no blanks can be given, then one test exposure with blanks. Blanks filled in with original text words would constitute a word-level verbatim score; those filled in with original or similar words in the same grammatical form class would be a word-level substance score.

Examples of verbatim, cued recall items would be:

1. The Crusades were expeditions to recover the ____ _____.
2. The Muslim who recaptured Jerusalem was _____.

Examples of substance cued recall items would be:

1. From the beginning of the First Crusade to the end of the Fourth spanned _____years.
2. The Christians recovered Jerusalem in the _____Crusade.

Some verbatim, multiple-choice items would be:

1. The Christians tried to recapture the Holy Land from the:

    a. Buddhists
    b. Hindus
    c. Muslims
    d. Druids

2. Saladin recaptured Jerusalem in:

    a. 1095
    b. 1146
    c. 1187
    d. 1200

Some substance, multiple-choice items would be:

1. The Crusades occurred about how many years ago?

    a. 500
    b. 800
    c. 1100
    d. 1400

2. The main reason for the Crusades was:

    a. criminal
    b. literary
    c. economic
    d. religious

Free recall, cloze, cued recall and recognition tests may offer a broad view of verbatim and substance learning. Such different response measures may also present different pictures of verbatim and substance information gain. As noted before, that verbatim learning was easier in the English et al. study and more difficult in the Cofer study may be due in part to the use of a recognition test in the former and a recall test in the latter. Yet another consideration is that of

adequate control conditions to determine if recognition and recall performance reflect actual gain in information as a result of passage presentation.

Since information processing is studied through examination of response measures, it is important to deal with these indices in detail. Therefore, the characteristics of the different response measures, the relationships between them, and the use of appropriate control conditions will be discussed in the following section.

III. MEASURES OF INFORMATION GAIN

In this section the measurement of comprehension in terms of gain of new information is considered. A general procedure for such measurement is proposed, and test methods based on that procedure are discussed. Also considered will be the memory requirements of the various test methods.

A general method for testing information gain. To measure how much new information a student has acquired after exposure to the reading stimulus, it is necessary to have a pre-measure of how much relevant information the student already has acquired prior to stimulus presentation (Marks & Noll, 1967). Therefore it is necessary to use a pre- and posttest procedure regardless of the different types of tests which may be used to assess comprehension. The difference between pretest and posttest performance will be defined as information gain. However, it is clear that the difference in performance between pretest and posttest may be a function of variables other than the stimulus presentation itself. For example, the pretest may operate to cue the student as to what information in the passage is relevant. This may result in higher posttest scores. It is also possible that the pretest could fixate responses to test items and thus interfere with posttest performance. To estimate the effects of the pretest on posttest performance and to isolate the factor of previously attained information, a series of experiments will be undertaken which, by use of appropriate controls, will provide a relatively good estimate of new information gain.

Evidence for comprehension of connected discourse can be gathered by one of several recall and recognition methods. For all methods, a passage is first presented and later the reader is asked to demonstrate that he has learned something as a result of reading. There are two notable exceptions to this generalization: (1) the cloze procedure in which the student is given a passage with every n^{th} word deleted (without prior exposure to the undeleted passage) and the student is asked to fill in the missing words; and (2) the class of procedures in which the passage is available at the time of testing. This allows the S to search the passage for the correct response so that recall requirements under this condition are minimal.

Four general classes of comprehension response measures have been identified: recognition, free recall, cued recall, and serial recall. The four are discussed below.

Recognition

Murdock (1963) identifies three different recognition tasks: (1) Tests which require a binary choice (True-False, Yes-No). In terms of a comprehension measure this method requires S to respond as to whether the information in a test item statement is congruent with the information in the test passage. (2) Tests which require the S to select a "correct" response from a set of alternatives in response to a question or statement concerning the passage. This type of task is best exemplified by the traditional multiple-choice test. (3) Tests which require the S to select all correct responses from a relatively large number of alternatives. In this situation there are a number of "correct" alternatives and S must attempt to find as many correct alternatives as possible. This method appears to be a combination of both (1) and (2).

Recognition responses may be measured by accuracy of response, latency, or both. Furthermore, as Murdock suggests, there appears to be no reason to believe that the three recognition tasks involve different processes, since all involve the selection of a correct response from a number of alternatives.

For the purposes of this paper, only the multiple-choice type task will be considered.

Some Variables Which Affect Multiple-Choice Responding

(1) Response Biases - Although a number of responses biases have been identified (Cronbach, 1946), only two will be considered here. The first is the tendency to guess when the correct alternative is not known. Ss tend to differ in their willingness to guess (Gritten & Johnson, 1941; Gilmour & Gray, 1942; Wood, 1926). This bias is most often handled by either encouraging all Ss to guess whenever they are in doubt or to discourage Ss from guessing. Ss who have a greater tendency to guess are likely to produce higher scores since by chance they will respond correctly on some items which are not responded to by Ss who do not guess. Intuitively it would appear that it would be easier to encourage all Ss to guess rather than to discourage all from guessing.

The second bias is position preference. The question here is whether Ss have a tendency to choose certain alternatives simply by their position in the set. For example, there seems to be some evidence that Ss prefer to guess alternatives (a) or (b) rather than (c) or (d) in a four-alternative situation (Gustav, 1963). This bias is usually handled by randomizing the position of the correct alternative. This,

however, rectifies the test constructor's bias rather than the testee's bias. That is, randomization will only insure that if biased guessing occurs the student will, on the average, not be correct more than the guessing probability would predict. The fact remains, however, that this position bias may result in poorer than chance performance.

(2) Number of response alternatives. The number of response alternatives may affect performance in two ways. First, as the number of response alternatives increases, performance may decrease because it is more difficult to discriminate the correct one. Second, increasing the number of response alternatives may make elimination of incorrect responses more difficult. In data collected by Murdock (1963) there is evidence that Ss use a strategy in which they attempt to elimate as many incorrect alternatives as possible and guess randomly from the remaining alternatives. He demonstrates that as the number of alternatives (2,3 or 4) increases, and scores are corrected for guessing, performance in terms of mean number correct decreases. As Murdock points out, it is important to distinguish between items in which the S knows the correct alternative and responds accordingly and items in which the S does not know or is unsure of the correct alternative and eliminates some responses and chooses randomly from the set of remaining alternatives.

(3) Quality of distractors. It is obvious that the discriminability of test items in a multiple-choice recognition task will be a function of the quality of the distractor alternatives. If the distractors are obviously incorrect, then the task for the S becomes much simpler. Consider this item: Columbus discovered America in (a) 1936, (b) 1492 (c) 1894. If the alternatives were 1492, 1493, and 1494, the item would be more difficult. In the first set of alternatives the S can be correct simply by eliminating (a) and (c) whether or not he knows that the correct date is 1492. The correct response in the second set of alternatives is less obvious and S must not only know that America was discovered prior to 1894 but must know the exact year. Ideally the distractor alternatives should be such that if the S does not know the correct response all alternatives have an equal probability of being chosen.

Free Recall

In free recall tests the S is presented with a stimulus (a passage) and at some later time is asked to recall anything and everything that he can about the original stimulus. No constraints are imposed on the order in which information is to be recalled. As Deese & Hulse (1967) point out, the major difference between recognition and free recall is the memory requirements in the two tasks. In free recall the correct response can be one of a very large set of alternative responses, as for example, free recall of a list of unrelated words. Here the response set is the entire vocabulary of the language. In formal recognition tasks the number of alternatives is greatly delimited. We have already

discussed the fact that within recognition procedures, correctness is inversely related to the number of response alternatives. In free recall this effect is further exaggerated. Therefore, recognition will often show retention where free recall will not. However, it is not always the case that recognition is superior to free recall. Both Davis, Sutherland, & Judd (1961) and Erhlich, Flores, & LeNy (1960) have demonstrated that when the number of response alternatives controlled in free recall, retention is equal for recognition and free recall.

In terms of measuring information gain, free recall is a difficult procedure to use in the sense that construction of an appropriate pretest is difficult. Furthermore, since the E is usually interested in the gain of specific information (either factual or relational) the absence of that information in free recall indicates that the information is either not available for retrieval at that time or that the S does not think it important enough to report. Perhaps a more thorough test of recall would be a free recall test followed by a recognition test to tap unreported information so that both availability and recognition memory are tested.

Cued Recall

Cued recall refers to a modified recall situation in which the S is cued as to what it is he is expected to recall. The S is not supplied with alternatives but must generate his own response. This procedure is best exemplified by "fill-in" and "short-answer" questions (e.g., State "Boyle's Law." What were the names of Columbus' ships? Columbus discovered America in _____). The advantage of cued recall testing procedures over recognition testing procedures is that the former tests the availability of the information and is not subject to the confounding variables present in the multiple-choice task. The advantages of cued recall over free recall are that the cued recall measure does not depend solely on the S's disposition to give certain information and withhold other information, and that it gives a better measure of whether the S has a specific bit of information available. Further, pretest procedures for cued recall are easier to formulate than for free recall.

The problem with both free recall and cued recall is that the scoring is less objective than recognition scoring. Thus, for example, the scorer must decide what constitutes a correct statement of Boyle's Law. Both accuracy and latency of response are more difficult to compute than in recognition tasks.

Serial Recall

Serial recall refers to tests in which S is asked to recall information in the same sequentially organized fashion as given in the

text. Verbatim serial recall scores would note identical words, phrases and sentences in exactly the same position as in the text. Substance serial recall scores would cover synonymous words, phrases and sentences in relatively the same order as in the text.

The cloze procedure may be regarded as a combination of both serial and cued recall tasks. Three cloze procedures should be differentiated. In the most common case, S receives a passage with every n^{th} word deleted and fills in the missing words. A second procedure is one in which S first receives the undeleted version of the passage and subsequently receives the same passage with every n^{th} deleted. A third procedure is to present the deleted passage first, followed by the undeleted form, followed by a second presentation of the deleted form. This third method may provide a measure of information gain. A variation of this method was used by Coleman & Miller (1968). They used a guessing procedure similar to Shannon's method (1951) in which S is required to guess each word. After completing the passage, S goes through the same procedure again. The difference in correct "guesses" between the two trials is taken as a measure of information gain. It is clear that the cloze procedure is sensitive to S's ability to make use of context and linguistic redundancy but the usefulness of the cloze in the study of the comprehension process depends in part on the particular deletions and method of scoring. In measuring substance learning, content words will usually be of greater interest than the function words. Scoring synonyms will be more useful than scoring only exact-word responses.

Summary

A general method for measuring new information gain and various test procedures which might be employed in such measurement have been discussed. Recognition and cued recall appear to present the fewest problems in terms of measurement. Recognition and free recall procedures differentiate between discriminability of the correct response from a set of alternatives and the availability of the response for retrieval from memory. Since both of these processes are legitimate and important aspects of a comprehension curriculum, both will be dealt with in the author's planned research program.

The research program will investigate the several test methods outlined above and will attempt to devise information gain measures appropriate to each. Since in the authors' view comprehension is not a unitary process, each test measure devised is envisioned as a measure of a sub-process of comprehension, each tapping a separate but overlapping component of the overall construct.

IV. STIMULUS CHARACTERISTICS

Given a certain passage, a certain level of information gain may result. If the passage is altered in some way, then the level of in-

formation gain may also change. In this section the identification and manipulation of passage characteristics which can affect degree of comprehension are considered.

The selection of stimulus characteristics for investigation is influenced by the general approach toward comprehension which has been adopted. Recall that comprehension is viewed here as an information processing event, a complex of cognitive and learning operations. Moreover, the information processed is the meaning of connected discourse messages. Processing the meaning implies a central operation like organization. Consequently, stimulus characteristics thought to influence the organization of connected discourse will be of primary interest in the research program.

While not unimportant, factors such as word length and word abstractness have been extensively studied in readability research. Also, there is a growing body of literature on the effects of grammatical structure on comprehension. Although this area is important, it is outside the domain of this paper. Instead, other variables which have been examined less extensively, if at all, and which are on a less molecular level, are explored here. Two stimulus characteristics thought related to organization, topicality and sequential organization will be discussed.

Topicality

Topicality refers to the organization of the passage around a theme, to how tightly a passage is structured around the subject-matter. It does not refer to the effects upon comprehension of different subject matters, e.g., science, history, or music. It does not refer to the timeliness of the subject-matter.

Although associative and connotative indices of meaning were judged inadequate measures of substance learning, the concept of associative and connotative meaning can suggest methods of attacking topicality. For example, if the meaning of a word lies in its associations, if the meaning of a word is the structure of words surrounding it, then the meaning of a word might be transmitted by presenting its associations. Moreover, the stronger the associations, the tighter the organization and the more efficient the meaning transmission. Thus, given a passage about weather, greater organization and thereby greater topicality will result if strong associations of the word weather are included in the passage.

That organization can occur through associations is shown in clustering experiments. An implicit assumption is that some generalizations can be drawn from free recall of single words to recall of connected discourse. Bousfield (1953) presented a 60-item list of animals, male names, professions, and vegetables in random order. Subjects later recalled the items in category clusters. But more important, Bousfield, Cohen and Whitmarsh (1958) found that taxonomic clustering

was greater when the items were high-associative instances of the category names. For instance, Animal words such as dog and cat would lead to greater clustering than would ant and lion.

Furthermore, retention of category material shows stability. In Mandler's series of studies (1967), a list of 52 or 100 unrelated words was presented and subjects were asked to sort the words into different categories of their own choosing. After a consistent sorting system developed, subjects were asked to recall the list words. The mean percentage of items recalled across the Mandler series was about 46%. With original recall levels as baseline, the delayed recall scores at 3-4 days averaged 50% and went down to a stable 20-30% in from 3-15 weeks. Also, the clustering scores for lists remained above chance levels even after 14 weeks.

Mandler's results indicate that retention of organized material remains at a consistent level and does not fall to zero, and Bousfield's work suggests that taxonomic material prompts organizational activity, especially when associative relations are strong.

The effects of word association on connected discourse have been studied, but only the retention of the S-R word pairs has been examined, not the retention of the whole passage. Rosenberg (1966) embedded either 16 high- or 16 low-associative words in passages. Samuels (1968) tested both 5th- and 6th-grade and college students on the Rosenberg material. Both subject groups read the high-associative passage faster and both had higher scores on a 12-item multiple-choice test evaluating high-associative passage retention than on a comparable test evaluating retention of low-associative passage information. Samuels' multiple-choice test, like Rosenberg's, tapped only the S-R words; the correct alternative was a high-associative response to the stimulus word in the question for each test item.

The Rosenberg and Samuels studies did not measure retention beyond the S-R words, and therefore did not demonstrate the power of word associations as mnemonic devices for retention and organization of connected discourse. There was no evidence for a facilitating effect of word association on comprehension of the passage as a whole.

In investigating the associative aspects of connected discourse, one should go beyond retention of the isolated S-R pairs and examine retention and organization of the whole passage. Moreover, association should go beyond presentation of discrete S-R pairs and deal with mutual associative overlap or associative environments, where several words tend to evoke each other and thereby share common meaning (Deese, 1966).

Another way of analyzing topicality is through anaphoric analysis. Anaphora is the device for referring to an antecedent idea, which usually is in noun form. For instance, in the sequence "Jim hit the ball, then he ran," "he" is the anaphoric expression for "Jim." As the number of words between the anaphoric term and the original concept

increases, it would be expected that structuring around the original concept word would be more difficult. An anaphoric scoring method has been presented by Menzel (1968).

Initially, methods of scaling topicality in passages, possibly at the sentence level, will be sought. It is expected, however, that topicality of a passage sentence will depend a great deal on its position in the passage, on its relation to sentences which precede and follow it. This position variable is discussed further in the next section on sequential organization.

Sequential Organization

Sequential organization refers to the order of ideas in a passage. It is possible to distinguish between (1) sequential organization of ideas in terms of sentence syntax, and (2) sequential organization in terms of the logical or semantic order of sentences, regardless of the syntax of the individual sentences. While English syntax permits the expression of an idea in a number of syntactic orders, and the particular form of expression affects level of comprehension, this section is primarily concerned with the effect of sequential organization on an inter-sentence level. While the effects of intra-sentence syntactic order must constitute a significant aspect in a systematic approach to connected discourse comprehension, a detailed discussion of sentence syntax is beyond the scope of this paper.

Order effects in connected discourse have been studied in terms of serial learning, which requires memorization of a set of unrelated items in a specified order. Generally a bowed error curve has been found, reflecting greater difficulty in recalling items just past the middle of the list (Kausler, 1966, pp. 14-17). Deese and Kaufman (1957) presented 100-word passages consisting of 10 statements to college students. Passage topics were "Montana," "The Museum of Science and Industry," or "Bonneville Dam." For each passage, different arrangements of the 10 statements were given to different students. In recalling the passage after one presentation, subjects showed the classic serial position curve. Recall was lower for statements in the middle of the passage, and lowest just past the middle. Thus serial learning of unrelated items like nonsense syllables approximated memory for connected discourse.

Rothkopf (1962), however, failed to replicate Deese and Kaufman's results. He tested telephone operators and clerks on 12-sentence passages which dealt with a fictional primitive tribe, or Westminster Abbey, or a fictional European city. Once again sentence order within passages was counterbalanced across different subjects. Instead of a free recall test, subjects had 12 fill-in items, which were the 12 passage sentence with one deleted word. No serial position curve was found. Rothkopf concluded that the different test procedures (free recall vs. fill-in) probably accounted for the different outcomes.

The passages used in the Deese and Kaufman and the Rothkopf studies had mixed sentence order across subjects. This counterbalancing indicates that no logical sequence of ideas existed in the excerpts, that presenting a specific sentence order was not important to the sense of the passage. Yet in much connected discourse, there is a temporal or logical flow of ideas and the sentence order cannot be manipulated at will without destroying the meaning of the passage. Consider the following three sentences: "Washington was President in 1779. Jackson was President in 1829. And Lincoln was President in 1861." The temporal order of these sentences would be disrupted if changed to: "Jackson was President in 1829. Washington was President in 1779. And Lincoln was President in 1861." Or consider the logical sequence of "Jack ate too much. Then he felt sick." These sentences would make less sense as "Jack felt sick. Then he ate too much."

Comprehension differences by changes in logical order of sentences were found by Darnell (1960) with college undergraduates. A passage on readability had 15 sentences: a thesis sentence, two major contentions, two subcontentions for each contention, and two assertions for each subcontention. Seven sentence orders were used and a cloze test was given, with subjects filling in 48 blanks in the passage. Differences in cloze scores were obtained for the seven sentence orders, and performance was highest when a deductive order was given--thesis to contention to subcontention to assertion.

Thompson (1967) presented speeches which were prejudged as showing very high, high, moderate or low structure to college undergraduates rated high, medium or low in organizational ability. A 30-item multiple-choice test was given and all subjects, regardless of organizational ability, scored higher as speech structure increased. The difference between the "very high" and "high" structure speech was that the former included transitions which set the stage for, or summarized, the main points covered.

In the authors' planned investigation of sequential organization, both position effects and logical ordering in connected discourse will be investigated. It is expected that both sequential variables will be important, with degree of importance dependent on type of passage material.

V. STUDENT VARIABLES

In consonance with the Southwest Regional Laboratory's current focus of interest, the population of interest to this research program will be students from kindergarten through 6th grade. Further, since reading comprehension, not listening comprehension, is the primary concern, the target population consists of elementary grade subjects who have acquired the basic identification skills of reading.

Although many student variables can be identified (e.g., sex, personality, motivation, perceptual development, socioeconomic status), a

systematic investigation of these variables will not be undertaken at this time. Student variables, as they affect comprehension, are best studied in relation to the processes involved in comprehension. When more is known about the processes involved in comprehension, it will be possible to study how student variables relate to and affect the comprehension processes.

There is no attempt here to minimize the importance of student variables. Rather, the view here is that the nature of the problem and the state of knowledge about it are such that investigation of student variables must await the resolution of more fundamental problems.

VI. CONCEPTUAL ISSUES

The present approach to comprehension has been discussed in terms of attempting to construct a model or family of models for information processing of connected discourse. To this end, use had been made of various hypothetical constructs and of theory and research outside the field of "reading." The view is taken here that theory or model building serves two important functions: (a) to organize a wide range of experimental data under a common set of principles, and (b) to generate new, promising lines of investigation. Hopefully, a model of information processing of connected discourse will evolve which serves both functions of theory. In addition, it is likely that many of the constructs and much of the research on verbal learning and memory, while dealing mainly with single stimuli, are relevant to the study of connected discourse. For example, in the section on Stimulus Characteristics it was suggested that associative meaning offers a means of investigating connected discourse topicality. Apart from an interest in theory or models, an interest in verbal learning would lead one to think in terms of general explanations which go beneath surface results and beyond one or two test situations, and to consider elemental processes from which comprehension of connected discourse can be predicted and controlled.

One may question why psychological concepts are stressed over linquistic and educational ones. Unlike much linguistic work, the research in human learning and memory is empirical and tied to a system of experimental verifiability. Linguisitics, on the other hand, emphasizes taxonomic models of language competence and innate predispositions, whereas our concern is with performance models of actual language processing and adaptable learning. Educational concepts are not chosen since they often have no clear and precise operational definitions, or, if they do, are founded on correlational constructs which have little relation to processes.

Thus far, the discussion has focused mainly on definitions, types of information, information gain, stimulus and response variables which may affect information processing, and student variables. There has been little discussion of theoretical issues underlying information processes. The study of information processing raises various conceptual or

theoretical issues which relate to construction of a model of comprehension and thus warrant further discussion. The number of such issues exhausts the domain of human learning and memory. No attempt will be made to deal with all such issues here. Rather, the focus will be upon a subset of those issues dealing with memory functions. Since information processing is usually measured after reading in a recall or recognition test rather than concurrently with reading, retention is measured rather than actual ongoing comprehension activity. The remainder of this section will consider information storage and retrieval in terms of units of information, long-term and short-term memory, memory trace and interference theories of forgetting, and serial position effects in both stimulus and test item presentation order.

Short and Long-Term Memory

As noted earlier, information gain is examined in terms of memory because measures ot comprehension are typically taken after, rather than concurrently with, reading. Consequently, retention is measured instead of actual, ongoing processing of information. The retention usually examined is of the short-term variety, since comprehension tests are typically given immediately after passage presentation. However, long-term retention is at least as critical in any educational curriculum.

Short-term and long-term memory (STM and LTM) can be distinguished on the basis of: (a) The interval between stimulus presentation and retention test. Typically STM intervals range from less than one second to several minutes. LTM intervals have typically ranged from several minutes to several years. (b) Different memory processes, i.e., a transitory, limited storage capacity for STM and a more permanent, unlimited storage capacity for LTM. Prevalent theories to account for these process differences will be discussed in a subsequent portion of this paper. Although distinction (a) is imprecise, it serves as a useful distinction in comparing studies on retention or forgetting in terms of STM and LTM. In the following review of the effect of the length of the interval between stimulus presentation and the test recall no assumptions will be made concerning the upper-bound of STM or the lower-bound of LTM.

Although there has been a great deal of research on STM and LTM over the past decade, the following review will deal primarily with studies concerned with retention of information from connected discourse.

In the previously noted study of English, Welborn & Killian (1934), test passages were read and then true-false recognition tests were administered, with tests including items taken directly from the text and others being paraphrases or summaries of passage material. Verbatim scores usually dropped and substance scores usually rose as retention interval increased. For example, in the sixth experiment reported by English et al., Ss gave 63% correct verbatim response after 10 minutes'

delay and 57% at 30 days; correct substance responses were 35% after 10 minutes and 45% after 30 days.

The English et al. study was replicated by Briggs & Reid (1943) to check the reliability of high-substance LTM. College students read passages on educational psychology and then took a true-false recognition test. In contrast to the English et al. procedure, however, subjects were allowed unlimited reading time and independent groups were retested at different intervals following the immediate test for all groups. The five test-retest intervals were 0, 1, 4, 8, and 12 weeks. No verbatim retention levels were reported. Immediate substance retention scores centered around 70% for all groups, and scores dropped from 68% to 61% as retention interval increased from 1 to 12 weeks. While Briggs & Reid failed to replicate the rise in substance retention at long intervals, the substance scores were stable and higher than those reported by English et al.

Despite the differences between the English et al. and the Briggs & Reid studies in procedures and results, the high resistance to forgetting of substance information gain seems reliable. Yet substance retention does not remain high indefinitely. Cofer (1943) retested six subjects for recall four years after the original experimental session. Three of the six showed some recall, 11, 17 and 12%. Although low compared to the previous 50-70% recognition retention at 2-3 months, that any substance recall retention was found after four years appears remarkable.

A study of Dietze & Jones (1931) focused on verbatim LTM. High school students read 1000-1200-word passages which covered much factual material. These articles covered the discovery and uses of radium, the life of the early Germans, or a biography of the investor Sir Richard Arkwright. Multiple-choice recognition tests on specific facts were given immediately and 1, 14, 30 and 100 days after the reading. On all retention tests, scores increased as a function of grade (7th to 12th). The average immediate retention across grade level and passage topic was 64%, dropping to 35% at 30 days and 30% at 100 days. Comparing verbatim retention as reported in the English et al. and the Dietze & Jones studies, both reported initial retention of around 64%. After 30 days, however, English et al. reported a decrement of 7% while Dietze & Jones reported a decrement of 29%. The difference is verbatim LTM may be partly due to the use of a true-false test in the English et al. study and a multiple-choice test in the Dietze & Jones study.

Given that verbatim information gain declines from STM to LTM, and that substance gain is more durable than verbatim gain, the next step is discovering which factors influence retention levels, both STM and LTM, both verbatim and substance. In their sixth study, English, Welborn & Killian (1934) varied number of reading trials from one to four. As measured by the 10-minute test, both verbatim and substance retention increased as number of readings increased. For the 30-day test, verbatim retention increased with number of readings but substance

retention showed no differences with more readings. That is, a practice effect occurred for both verbatim and substance STM and verbatim LTM, but not for substance LTM. Substance LTM retention levels rose and converged, regardless of initial STM differences. Rothkopf (1968) studied practice effects in verbatim STM with the cloze procedure. Subjects took a cloze test after 0, 1, 2, or 4 exposures to the undeleted test passage. Correct responses increased as a negatively accelerated function of the number of passage exposures.

Another variable investigated in connected discourse retention is the material interpolated between the reading and test periods. McGeoch & McKinney (1934) examined retroactive inhibition in retention of passages on psychological testing. The test questions measured substance retention. In the interpolated period, the experimental group read a passage covering material highly related to the original passage, while the control group received a pitch discrimination test. Retention for original learning was tested before and after the interpolated period. Both experimental and control groups showed higher retention after the interpolated task than before it, 2 to 6% higher for experimental and 8 to 12% higher for control. Consequently, absolute retroactive effects were facilitative instead of inhibitory. However, control group superiority suggests that relatively greater retroactive inhibition characterized the experimental group.

Deese & Hardman (1954) and Hall (1955) investigated the effects of similarity between the original and interpolated material. Deese & Hardman calculated the number of content words recalled from the original passage, and found no significant retroactive inhibition in the experimental condition. Hall presented a fill-in test, passage sentences with a deleted word per sentence, and reported virtually no difference between experimental and control groups. Whereas McGeoch & McKinney apparently measured substance retention, both Deese & Hardman and Hall unquestionably assessed verbatim retention. Yet all three reported similar results: connected discourse is highly resistant to retroactive inhibition.

The McGeoch & McKinney and the Hall experiments retested students a few days later. With immediate retention scores as the baseline, McGeoch & McKinney reported a slight gain in the control condition and a slight loss in the experimental at 7 days. Hall found about a 15% retention drop at 21 days, and almost no difference between experimental and control groups. Therefore, McGeoch & McKinney's substance retention measure after one week indicated high LTM and little retroactive inhibition, and Hall's verbatim measure at three weeks showed a slight decrement in LTM and no retroactive inhibition.

So far, it appears that connected discourse is highly resistant to interference by interpolated material. In a series of studies, Slamecka (1959, 1960a, 1960b, 1962) investigated whether procedural variations could prompt retroactive inhibition in connected discourse. In these studies, 20-18-word, single-sentence passages were used and strict

verbatim retention was measured (number of correct words in the correct positions).

Slamecka (1959) varied similarity of interpolated material to the original passage and found that recall decreased as similarity increased and that the high, medium and low similarity groups all had lower retention than the control group which had no intervening passage. Two other studies in the same report varied number of presentation trials for the original passage and the interpolated passage. Recall increased with increasing original learning and with decreasing interpolated learning.

Slamecka (1960a) again varied degree of original and interpolated learning, but in one factorial design combining 2, 4, or 8 original learning trials and 0, 4 or 8 interpolated learning trials. Again, recall increased with additional original learning trials and decreased with additional interpolated learning trials.

Slamecka (1960b) studied the effects of similarity between the original and interpolated passages. Rankings of high, medium and low similarity were determined pre-experimentally by six judges. As before, retention decreased as similarity increased, and all three similarity groups showed less retention than the control group with no interpolated passage.

Slamecka (1962) presented three experiments which varied the interpolated task. The first study used 0, 6, 12 or 18 interpolated passages, the second used 0, 2, 4 or 6 different passages (each presented three times), and the third used 0, 3, 6 or 9 interpolated trials (each of which was double the length of the passages in the first study). In all three cases, retention decreased as amount of interpolated material increased. After comparing the results across the three studies, Slamecka concluded that the degree of retroactive inhibition depended on the duration of the interpolated learning session.

In summarizing the research on verbatim and substance STM and LTM, two conclusions seem warranted. First, substance LTM is much stronger than verbatim LTM. Second, verbatim STM and LTM are more sensitive to experimental manipulation, such as practice level and type of interpolated material, than substance STM and LTM. Both conclusions are disturbing. If retention of substance information gain is resistant to variables found effective in manipulating retention of unrelated verbal stimuli, then the findings from much verbal learning research cannot be generalized to connected discourse. If retention of substance gain shows little forgetting, perhaps existing explanations for forgetting do not apply to connected discourse.

Whether verbal learning research and theory should be rejected depends on further connected discourse research. Moreover, STM and LTM research and theory should be reviewed so that it is clear what one is accepting or rejecting. The next section will consider STM

and LTM research in the light of two theories of forgetting, interference and decay.

Interference and Memory Trace Theories

Much of the early work on retention, starting with Ebbinghaus (1885), has been in LTM. In recent years, interest and research in STM was generated by the now-classic study of Peterson & Peterson (1959) on single stimulus STM. In analyzing STM there is a tendency to take the uniprocess view and to apply LTM concepts to STM, as does Melton (1963). An alternative view proposes that STM involves different processes than LTM (Hebb, 1949). Interference theory which is drawn from LTM research will be discussed first, and then decay theory which is proposed for STM will be considered.

Interference theory declares that forgetting depends on interfering material which precedes or follows the learning items, i.e., proactive and retroactive inhibition (PI and RI). Such inhibition leads to forgetting through response competition and unlearning. Both processes can be noted in a study by Barnes & Underwood (1959). An A-B, A-C transfer paradigm was used: a second list of paired associates repaired stimuli from the first line with new responses. After the A-C learning, subjects were asked to recall both the first- and second-list responses. Subjects given more A-C trials recalled fewer first-list B responses (unlearning), and when both responses were available, the second-list responses were recalled first (response competition).

During the last 10 years, PI has been considered more important than RI in forgetting. Underwood (1957) reported that the amount of RI reported in previous experiments was a direct function of the number of prior lists Ss had practiced before learning the test list. Underwood & Postman (1960) proposed that PI is produced not only by requiring repeated learning of lists in the laboratory, but also by extra-experimental linguistic habits which Ss bring to the laboratory situation. Two potential sources of extra-experimental interference were identified: letter-sequence and unit-sequence factors giving rise to the letter-letter and word-word associations which could interfere with learning new material in the laboratory. Postman (1961) discusses the letter- and unit-sequence research in greater detail.

More recently, however, Postman (1963) and Underwood & Exstrand (1966) have taken issue with the Underwood & Postman formulation, since it apparently fails to predict rates of forgetting. Keppel (1968) suggested that forgetting may be better accounted for by non-specific linguistic activity occurring during the retention interval which results in unlearning, rather than by emphasis on extra-experimental sources of PI during the acquisition stage.

Despite the differences in the identification and emphasis of interference sources, forgetting in interference theory remains a function of

two basic processes, RI and PI. But the adequacy of RI and PI explanations, and thereby of interference theory, has been questioned as a result of STM findings.

Memory trace theory (Hebb, 1949) proposes that neural traces associated with stimulus presentations decay quickly and consequently render the stimulus unavailable for recall in STM. If repeated stimulus presentations occur, then permanent traces are established for LTM. The critical factor in trace theory is the delay interval between stimulus presentation and recall.

Peterson & Peterson (1959) showed trigrams for brief intervals, followed by an interval of 3 to 18 seconds during which Ss counted backwards. Retention tests revealed rapid forgetting with scores dropping to 10% of initial retention after 18 seconds' delay. Peterson (1963) favored a memory trace explanation because the interpolated number-counting was different from the verbal memory task (no RI) and because no retention loss was noted when subjects received repeated stimulus presentations (no PI). However, Postman (1964) argued that a supposedly irrelevant interpolated task could offer "generalized response competition" and produce RI. Based upon a finer analysis of Peterson's and their own data, Keppel & Underwood (1962) suggest that PI effects which developed early were soon offset by practice effects, although the data provide no direct evidence for this contention.

A more direct test of the memory trace theory was carried out by Waugh & Norman (1965). A list of 16 single digits was read and Ss were cued to recall one digit in the series. The independent variables were position of the test digit in the series and rate of digit presentation. If RI is the critical factor, the number of digits following the test digit (the position of the test digit in the series) would determine amount of retention. If memory trace is the important factor, the time between the test digit and the recall test (the rate of digit presentation) would determine amount of recall. Results indicated that position of the test digit in the series was the main determinant of recall.

This study, in relation to the others already discussed, seems to indicate that forgetting in STM is a function of both trace decay and interference effects. However, it is possible that the effect of interference in STM is caused by a different process than in LTM. In LTM the effect appears to be primarily one of causing unlearning and S-R confusion while in STM the effect of previous stimuli may result in overloading of the STM storage capability.

While Peterson (1966 a,b) accepts interference as the major factor in STM, memory trace is still considered a factor of some importance. Also, establishing interference as the overriding mechanism for forgetting does not mean that STM and LTM can be lumped together in a single memory system. Waugh & Norman (1965) suggested a primary memory

store which parallels STM and a secondary memory store which covers LTM. Similarly, Atkinson & Shiffrin (1967) hypothesize separate short- and long-term memory stores. Both the Waugh & Norman and Atkinson & Shiffrin formulations have attention as an important variable in determining which stimuli enter the short-term system, and rehearsal as critical in determining what information flows from the short- to the long-term system. Both theories also reflect the recent interest in memory storage, memory retrieval, and the processing of information from short- to long-term storage.

Whether these new developments in STM and LTM theory and research prove relevant to connected discourse is an empirical question. An important consideration in assessing the validity of the memory models is their ability to cope with functional stimuli which may differ from the nominal stimulus and with organizational activity which subjects might impose on unrelated stimuli. This problem of defining the units of information processing is discussed in the next section.

Units of Information Processing

The questions of concern with regard to units of processing may be stated as follows: What are the units of processing for storage of information from connected discourse? Do these units transcend the word level by some kind of "chunking" (Miller, 1956) or "clustering" (Jenkins, Mink & Russel, 1958) mechanism on the intra- or intersentence level? Are the units of storage identical to the units of retrieval? Are the units identical to the stimulus input or are there transformations applied to the information input? Is the size or type of unit invariant over stimulus material? If not, what are the variables which affect the type and size of the units of processing?

Although the answers to these questions are not available, certain evidence concerning clustering and chunking is relevant to a systematic approach to these questions.

Clustering refers to changes in the order of recall as organization is introduced into the stimulus material. Typically, investigation of clustering has been based on either free-association or conceptual categories. Free-association data is exemplified by the work of Jenkins and his associates (Jenkins & Russell, 1952; Jenkins, Mink & Russel, 1958) in which they found that in free recall of single words, Ss tend to recall words in pairs which are associated or which have a mediated common association. These pairs tend to be recalled together in proportion to the frequency occurrence in free-association norms.

Category clustering is best exemplified by the investigations of Bousfield (Bousfield, 1953; Bousfield, Cohen & Whitmarsh, 1958). In these studies lists of words are presented for free recall. The words fall into several conceptual categories but these categories are not

pointed out to the Ss. The result of these studies is that words from the same conceptual category tend to be recalled in a cluster. Bousfield argues that words that are related tend to facilitate recall by arousing a mediation mechanism; in this case a supraordinate concept. The studies on clustering do not provide clear evidence whether the effect is a function of word associations or conceptual categorization. It should be possible, however, to design an experiment using appropriate stimulus material which would shed further light on this matter. It should be possible, for example, to construct a mixed list in which some words are strong free-associates of each other but are not in the same conceptual category, and others that are in the same conceptual category but are not strong associates. Such experiments might indicate whether both types of clustering are independent or whether associative clustering alone can account for the clustering phenomenon. What role clustering may play in information processing from connected discourse can only be hypothesized. It is possible, for example, that in reading a passage certain common associates of the words or phrases and certain mediated conceptual categories are retrieved and encoded back to the original information in a verbatim or substance mode.

Miller (1956) argued that the human memory system has a fixed span of about seven "chunks" of information. Miller argued that the reason people can recall more individual words in high approximations to English than in lower orders of approximation is that people recode information in chunks. Sentences are not merely strings of individual words but rather are organized by certain grammatical rules and related by association and meaning. Tulving & Patkau (1962) demonstrated that when the unit of recall or chunk is an "unbroken sequence" rather than a single word, there is little change in the number of chunks recalled as a function of order of approximation to English. What did change, however, as a function of order of approximation to English was the size of the chunk. In zero order approximations (randomly chosen words) the chunk was typically one word. For higher orders of approximation the chunks consisted of several words.

Both clustering and chunking mechanisms are viewed as useful constructs and the initial approach to investigating units of information will attempt to exploit these constructs as far as possible. The planned study of "information units" will be concerned with the relationship between clustering and chunking as well as how children can be instructed in efficient methods of unitizing information for efficient storage and retrieval. The primary concern with substance learning makes it likely that neither clustering nor chunking will be adequate constructs for defining units of information storage and retrieval. It is likely that additional processes will be needed to explain the retrieval of the substance of the message and the longevity of substance learning compared to verbatim learning.

Serial Position Effects for Recall of Connected Discourse.

Two aspects of serial position effects are of interest: (1) The effect of serial position of information within a passage, and (2) Serial position of test items in the set of items. The effect of serial position of information within the passage on comprehension testing may provide some evidence on how information is processed, stored, and retrieved. Furthermore, the relationship between topicality and sequential organization on the one hand, discussed earlier, and serial position effects on the other hand may provide a useful measure of passage organization. The effect of serial position of test items will provide methodological evidence on test construction and evaluation of test results as well as evidence on retrieval processes.

The serial position effect as a function of method of recall and structure of the stimulus material will be considered. Using the method of free recall Murdock (1962) showed that for a list of unrelated English words, the items that had the greatest probability of recall were those presented at the end of the list, followed by the items presented at the beginning of the list. Those items toward the middle of the list had the lowest probability of being reported in free recall. Both Deese & Kaufman (1957) and Bousfield, Cohen & Silva (1956) demostrated that the most frequently recalled items in free recall are those that occur last in stimulus presentation and first in recall.

Using the free recall method, Postman and Phillips (1965) showed that as the interval between presentation of the list and recall increases, the recency effect decreases. The interpretation of this result in terms of memory storage models is that in immediate recall the items at the end of the list are still in short-term memory while earlier items have either been lost by trace decay or have gone into long-term storage. Delayed recall tests are tapping long-term memory and, therefore, items at the end of the list lose their advantage in recall. The primary effect is interpreted by positing that the short-term storage is empty at the beginning of the list and, therefore, there is a greater opportunity for rehearsal and less susceptibility to mutual interference among items at the beginning of the list.

In the serial anticipation method of recall the serial position effect is almost the mirror image of the effect in free recall. Under conditions of serial anticipation items at the beginning of the list have the highest probability of recall, followed by the items at the end of the list. Again, the items toward the middle of the list have the lowest probability of recall (Hovland, 1938; McCrary & Hunter, 1953). Since the structure of English text, as discussed earlier in this paper, typically has some degree of sequential organization, one might hypothesize that recall of connected discourse should more closely follow the serial position curve found for the serial anticipation method than for the free recall method, regardless of the recall test used. That is, the nature of the stimulus in connected discourse is

such that Ss impose a serial order of recall not unlike serial anticipation. Therefore, in connected discourse, information from the beginning of the passage should have a greater probability of recall than information from the end of the passage. A test of this hypothesis was carried out by Deese & Kaufman (1957). Since this study was discussed in an earlier section, it will not be discussed in detail here. Suffice it to say that Deese & Kaufman found supporting evidence for the hypothesis that the serial position effect from connected discourse, using a free recall method, approximates the effect found when the serial anticipation method is used in list learning. Rothkopf (1962), using different passages and cued recall rather than free recall, did not replicate the Deese & Kaufman results. He did, however, find a serial position effect as a function of the order of presentation of items in the test. This effect, though not as pronounced as the Deese & Kaufman results, showed a serial anticipation position effect. This may suggest that the observed serial position effect is a function of some retrieval process rather than a storage process. Also, since the two studies did not use the same passages, the possibility remains that the Rothkopf passages had less sequential organization than the Deese & Kaufman passages. Deese & Kaufman presented material from zero order approximations of English to actual English text. Since they found a rather continuous shift from the typical free recall serial position effect to the typical serial anticipation position effect as a function of increase in order of approximation to English, the difference in results obtained by Rothkopf might partially, at least, arise from differences in sequential organization in the stimulus material.

The conclusion to be drawn from these studies is that the effect of serial position of information in connected discourse has not been satisfactorily resolved. Since the proposed research program is concerned with information storage and retrieval, the serial position effect will be of substantial importance.

VII. SUMMARY

The purpose of this paper was not to describe a model or theory of comprehension but rather to review past efforts and to indicate the authors' general approach as well as outlining some promising areas of investigation. Clearly, readers of this paper will question the inclusion of some areas and exclusion of others. This conceptualization is not presented as the "last word." The conceptualization will hopefully grow, and will consequently be revised as events warrant revision.

REFERENCES

Atkinson, R.C. & Shiffrin, R.M. Human memory; a proposed system and its control processes. In K.W. Spence & J.T. Spence (Eds.), The Psychology of Learning and Motivation. Vol. 2. New York: Academic Press, 1968, 89-195.

Barnes, J.M. & Underwood, B.J. "Fate" of first-list associations in transfer theory. Journal of Experimental Psychology, 1959, 58, 97-105.

Bedell, R.C. The relationship between the ability to recall and the ability to infer in specific learning situations. Unpublished doctoral dissertation, University of Missouri, 1934.

Betts, E.A. Reading as a thinking process. High Points, 1956, 38, 17-25.

Bousfield, W.A. The occurrence of clustering in the recall of randomly arranged associates. Journal of General Psychology, 1953, 49, 229-240.

Bousfield, W.A. The problem of meaning in verbal learning. In C.N. Cofer (Ed.), Verbal Learning and Verbal Behavior. New York: McGraw-Hill, 1961, 89-91.

Bousfield, W.A., Cohen, B.H. & Silva, J.G. The extension of Marbe's law to the recall of stimulus words. American Journal of Psychology, 1956, 69, 429-433.

Bousfield, W.A., Cohen, B.H. & Whitmarsh, G.A. Associative clustering in the recall of different taxonomic frequencies of occurrence. Psychological Reports, 1958, 4, 39-44.

Briggs, L.J. & Reid, H.B. The curve of retention for substance material. Journal of Experimental Psychology, 1943, 32, 513-517.

Carroll, J.B. The analysis of reading instruction: Perspectives from psychology and linguistics. Yearbook of the National Society for the Study of Education, 1964, 63, 336-353.

Cleland, D.L. A construct of comprehension. In J.A. Figurel (Ed.), Reading and Inquiry: International Reading Association Conference Proceedings. Newark, Del.: International Reading Association 1965, 59-64.

Cofer, C.N. A comparison of logical and verbatim learning of prose passages of different length. American Journal of Psychology, 1941, 54, 1-20.

Cofer, C.N. Recall of verbal materials after a four-year interval. Journal of General Psychology, 1943, 29, 155-156.

Coleman, E.B. & Miller, G.R. A measure of information gained during prose learning. Reading Research Quarterly, 1968, 3, 369-386.

Cronbach, L.J. Response sets and test validity. Educational and Psychological Measurement, 1946, 6, 475-494.

Darnell, D.K. The relationship between sentence order and comprehension. Speech Monograph, 1963, 30, 97-100.

Davis, F.B. Fundamental factors in comprehension in reading. Psychometrika, 1944, 9, 185-197.

Davis, F.B. The teaching of comprehension in reading in the secondary school. Education, 1956, 76, 541-544.

Davis, F.B. The role of testing in reading instruction. Conference on Reading, University of Chicago, 1966, 28, 178-189.

Davis, R., Sutherland, N.S. & Judd, B.R. Information content in recognition and recall. Journal of Experimental Psychology, 1961, 61, 422-429.

Deese, J. The Structure of Associations in Language and Thought. Baltimore: The Johns Hopkins University Press, 1966.

Deese, J. & Hardman, G.W. An analysis of errors in retroactive inhibition of rote verbal learning. American Journal of Psychology, 1954, 67, 229-307.

Deese, J. & Hulse, S.H. The Psychology of Learning. New York: McGraw-Hill Book Co., 1967.

Deese, J. & Kaufman, R.A. Serial effects in recall of unorganized and sequentially organized verbal material. Journal of Experimental Psychology, 1957, 54, 180-187.

Dietze, A.G. & Jones, G.E. Factual memory of secondary school pupils for a short article which they read a single time. Journal of Educational Psychology, 1931, 22, 586-598.

Ehrlich, S., Flores, C. & LeNy, J. Rappel et reconnaissance d'elements appartenant a des ensembles definis. Annee Psychologique, 1960, 60, 29-37.

English, H.B., Welborn, E.L. & Killian, C.D. Studies in substance memorization. Journal of General Psychology, 1934, 11, 233-260.

Gans, R. A study of critical reading comprehension in the intermediate grades. Contributions to Education, No. 811, New York: Bureau of Publications, Teachers College, Columbia University, 1940.

Gilmour, W.A. & Gray, D.E. Guessing on true-false tests. Educational Research Bulletin, 1942, 21, 9-12.

Gray, W.S. Reading and understanding. Elementary English, 1951, 28, 148-159.

Gray, W.S. The major aspects of reading. In H.M. Robinson (Ed.) Sequential Development of Reading Abilities, Supplementary Educational Monographs, University of Chicago Press, 1960, 8-21.

Gritten, F. & Johnson, D.M. Individual differences in judging multiple-choice questions. Journal of Educational Psychology, 1941, 32, 423-430.

Guilford, J.P. Frontiers in thinking that teachers should know about. Reading Teacher, 1960, 13, 176-182.

Gustav, A. Response set in objective achievement tests. Journal of Psychology, 1963, 56, 421-427.

Hall, J.F. Retroactive inhibition in meaningful material. Journal of Educational Psychology, 1955, 46, 47-52.

Hebb, D.O. Organization of Behavior. New York: John Wiley & Sons, 1949.

Holmes, J.A. The substrata-factor theory of reading: Some experimental evidence. In J.A. Figurel (Ed.) New Frontiers in Reading. International Reading Association Proceedings, New York: Scholastic Magazine, 1960, 115-120.

Holmes, J.A. Speed, comprehension and power in reading. Eleventh Yearbook of the National Reading Conference. Milwaukee: National Reading Conference, Inc., 1962, 6-14.

Holmes, J.A. Basic assumptions underlying the substrata-factor theory. Reading Research Quarterly, 1965, 1, 5-28.

Hovland, C.I. Experimental studies in rote-learning theory: II. Reminiscence with varying speeds of syllable presentation. Journal of Experimental Psychology, 1938, 22, 338-353.

Jenkins, J.J., Mink, W.D. & Russell, W.A. Associative clustering as a function of verbal association strength. Psychological Reports, 1958, 4, 127-136.

Jenkins, J.J. & Russell, W.A. Associative clustering during recall. Journal of Abnormal and Social Psychology, 1952, 47, 818-821.

Jones, M.G. & English, H.B. Notional vs. rote memory. American Journal of Psychology, 1926, 37, 602-603.

Johnson, M.S. Factors in reading comprehension. Educational Administration and Supervision, 1949, 35, 385-406.

Jordan, L.V. Building the readiness program. Mental Retardation, 1967, 5, 28-31.

Kausler, D.H. Readings in Verbal Learning: Contemporary Theory and Research. New York: John Wiley & Sons, 1966.

Keppel, G. Retroactive and proactive inhibition. In T.R. Dixon & D.L. Horton (Eds.), Verbal Behavior and General Behavior Theory. Englewood Cliffs, New Jersey: Prentice-Hall, 1968, 172-213.

Keppel, G. & Underwood, B.J. Proactive inhibition in short-term retention of single items. Journal of Verbal Learning and Verbal Behavior, 1962, 1, 153-161.

Kingston, A.J. A conceptual model of reading comprehension. Tenth Yearbook of the National Reading Conference. Milwaukee: National Reading Conference, Inc., 1961, 100-107.

Kingston, A.J. Some thoughts on reading comprehension. Eleventh Yearbook of the National Reading Conference. Milwaukee: National Reading Conference, Inc., 1962, 20-23.

Langsam, R.S. A factorial analysis of reading ability. Unpublished doctoral dissertation, New York University, 1941.

McCrary, J.W., Jr. & Hunter, W.S. Serial position curves in verbal learning. Science, 1953, 117, 131-134.

McCreary, A.P. & Surkan, A.J. The human reading process and information channels of communication systems. Journal of Reading, 1965, 8, 363-372.

McGeoch, J.A. & McKinney, F. The susceptibility of prose to retroactive inhibition. American Journal of Psychology, 1934, 46, 429-436.

McKee, P. The Teaching of Reading in the Elementary School. Boston: Houghton Mifflin Co., 1948.

Mandler, G. Organization and memory. In K.W. Spence & J.T. Spence (Eds.), The Psychology of Learning and Motivation. Vol. 1 New York: Academic Press, 1967, 327-372.

Marks, E. & Noll, G.A. Procedures and criteria for evaluating reading and listening comprehension tests. Educational and Psychological Measurement, 1967, 27, 335-348.

Melton, A.W. Implications of short-term memory for a general theory of memory. Journal of Verbal Learning and Verbal Behavior, 1963, 2, 1-21.

Menzel, P. Anaphora analysis. SWRL Supporting Paper, 1968.

Miller, G.A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review, 1956, 63, 81-97.

Murdock, B.B. The serial position effect of free recall. Journal of Experimental Psychology, 1962, 64, 482-488.

Murdock, B.B. An analysis of the recognition process. In C.N. Cofer & B.S. Musgrave (Eds.), Verbal Behavior and Learning: Problems and Processes. New York: McGraw-Hill Book Co., 1963.

Niles, O.S. Comprehension skills. Reading Teacher, 1963, 17, 2-7.

Nordberg, R.B. Levels of communication in reading. Catholic Educational Review, 1956, 54, 92-100.

Osgood, C.E. Comments on Professor Bousfield's paper. In C.N. Cofer (Ed.), Verbal Learning and Verbal Behavior. New York: McGraw-Hill, 1961, 91-106.

Osgood, C.E., Suci, G.J. & Tanenbaum, P.H. The Measurement of Meaning. Urbana, Illinois: University of Illinois Press, 1957.

Park, W. Reading or studying! What's the difference? Education Digest, 1966, 32, 49-51.

Peterson, L.R. Immediate memory: Date and theory. In C.N. Cofer & B.S. Musgrave (Eds.), Verbal Learning and Verbal Behavior, New York: McGraw-Hill, 1963, 336-370.

Peterson, L.R. Short-term memory. Scientific American, 1966, 215, 90-95. (a)

Peterson, L.R. Short-term verbal memory and learning. Psychological Review, 1966, 73, 193-207. (b)

Peterson, L.R. & Peterson, M.J. Short-term retention of individual verbal items. Journal of Experimental Psychology, 1959, 58, 193-198.

Postman, L. The present status of interference theory. In C.N. Cofer (Ed.), Verbal Learning and Verbal Behavior. New York: McGraw-Hill, 1961, 152-196.

Postman, L. Does interference theory predict too much forgetting? Journal of Verbal Learning and Verbal Behavior, 1963, 2, 40-48.

Postman, L. Short-term memory and incidental learning. In A.W. Melton (Ed.), Categories of Human Learning. New York: Academic Press, 1964, 145-201.

Postman, L. & Phillips, L.W. Short-term temporal changes in free recall. Quarterly Journal of Experimental Psychology, 1965, 17, 132-138.

Robinson, H.M. Development of reading skills. Elementary School Journal, 1958, 58, 269-274.

Robinson, H.M. The major aspects of reading. Conference on Reading, University of Chicago, 1966, 28, 22-32.

Rosenberg, S. Associative factors in the recall of connected discourse. Psychonomic Science, 1966, 4, 53-54.

Rothkopf, E.Z. Learning from written sentences: Effects of order of presentation on retention. Psychological Reports, 1962, 10, 667-674.

Rothkopf, E.Z. Textual constraint as function of repeated inspection. Journal of Educational Psychology, 1968, 59, 20-25.

Sachs, J.A. Recognition memory for syntactic and semantic aspects of connected discourse. Perception & Psychophysics, 1967, 2, 437-442.

Samuels, S.J. Effect of word associations on reading speed, recall, and guessing behavior on tests. Journal of Educational Psychology, 1968, 59, 12-15.

Semmelroth, C. Reading as information processing. North Central Reading Association Yearbook, 1965, 3, 85-93.

Shannon, C.E. A mathematical theory of communication. Bell System Technical Journal, 1948, 27, 349-423, 623-656.

Shannon, C.E. Prediction and entropy of printed English. Bell System Technical Journal, 1951, 30, 50-64.

Shannon, C.E. & Weaver, W. The Mathematical Theory of Communication. Urbana, Illinois: University of Illinois Press, 1949.

Simmons, J.S. Reasoning through reading. Journal of Reading, 1965, 8, 311-314.

Slamecka, N.J. Studies of retention of connected discourse. American Journal of Psychology, 1959, 72, 409-416.

Slamecka, N.J. Retroactive inhibition of connected discourse as a function of practice level. Journal of Experimental Psychology, 1960, 59, 104-108. (a)

Slamecka, N.J. Retroactive inhibition of connected discourse as a function of similarity of topic. Journal of Experimental Psychology, 1960, 60, 245-249. (b)

Slamecka, N.J. Retention of connected discourse as a function of duration of interpolated learning. Journal of Experimental Psychology, 1962, 63, 480-486.

Smith, D.E.P. Reading comprehension: A proposed model. In O.S. Cousey & E. Bliesmer (Eds.), Ninth Yearbook of the National Reading Conference for Colleges and Adults. Fort Worth, Texas: Texas Christian University Press, 1960, 21-27.

Spache, G.D. What is comprehension? Eleventh Yearbook of the National Reading Conference. Milwaukee: National Reading Conference Inc., 1962, 17-19.

Staiger, R.C. & Bleismer, E.P. Reading comprehension in the high school. Education, 1956, 76, 563-567.

Stauffer, R.G. Reading as a cognitive process. Elementary English, 1967, 44, 343-348.

Thompson, E. Some effects of message structure on listener's comprehension. Speech Monographs, 1967, 34, 51-79.

Thorndike, E.L. Reading as meaning: A study of mistakes in paragraph reading. Journal of Educational Psychology, 1917, 8, 323-332.

Traxler, A.E. Critical survey of tests for identifying difficulties in interpreting what is read. In W.S. Gray (Ed.), Promoting Growth Toward Maturity in Interpreting What is Read. Supplementary Educational Monographs, No. 74, University of Chicago Press, 1951, 195-200.

Trow, W.C. Recall vs. repetition in learning of rote and meaningful material. American Journal of Psychology, 1928, 40, 112-116.

Tulving, E. & Patkau, J.E. Concurrent effects of contextual constraint and word frequency on immediate recall and learning of verbal material. Canadian Journal of Psychology, 1962, 16, 83-95.

Underwood, B.J. Interference and forgetting. Psychological Review, 1957, 64, 49-60.

Underwood, B.J. & Ekstrand, B.R. An analysis of some shortcomings in the interference theory of forgetting. Psychological Review, 1966, 73, 540-549.

Underwood, B.J. & Postman, L. Extra experimental sources of interference in forgetting. Psychological Review, 1960, 67, 73-95.

Vernon, M.D. Learning and understanding. Quarterly Journal of Experimental Psychology, 1951, 3, 19-23.

Walcutt, C.C. Reading: A professional definition. Elementary School Journal, 1967, 67, 363-365.

Waugh, N.C. & Norman, D.A. Primary memory. Psychological Review, 1965, 72, 89-104.

Weaver, C.H. & Weaver, G.L. Information theory and the measurement of meaning. Speech Monographs, 1965, 32, 435-447.

Wiener, M. & Cromer, W. Reading and reading difficulty: A conceptual analysis. Harvard Educational Review, 1967, 37, 620-643.

Wood, B.D. Studies in achievement tests. Journal of Educational Psychology, 1926, 17, 1-22.

Yavuz, H.S. The retention of incidentally learned connotative responses. Journal of Psychology, 1963, 55, 409-418.